



ARTICLE

Learning Temporal User Features for Repost Prediction with Large Language Models

Wu-Jiu Sun¹ and Xiao Fan Liu^{1,2,*}

¹School of Computer Science and Engineering, Southeast University, Nanjing, 211189, China

²Department of Media and Communication, City University of Hong Kong, Hong Kong, 999077, China

*Corresponding Author: Xiao Fan Liu. Email: xf.liu@cityu.edu.hk

Received: 10 October 2024; Accepted: 23 December 2024; Published: 06 March 2025

ABSTRACT: Predicting information dissemination on social media, specifically users' reposting behavior, is crucial for applications such as advertising campaigns. Conventional methods use deep neural networks to make predictions based on features related to user topic interests and social preferences. However, these models frequently fail to account for the difficulties arising from limited training data and model size, which restrict their capacity to learn and capture the intricate patterns within microblogging data. To overcome this limitation, we introduce a novel model Adapt pre-trained Large Language model for Reposting Prediction (ALL-RP), which incorporates two key steps: (1) extracting features from post content and social interactions using a large language model with extensive parameters and trained on a vast corpus, and (2) performing semantic and temporal adaptation to transfer the large language model's knowledge of natural language, vision, and graph structures to reposting prediction tasks. Specifically, the temporal adapter in the ALL-RP model captures multi-dimensional temporal information from evolving patterns of user topic interests and social preferences, thereby providing a more realistic reflection of user attributes. Additionally, to enhance the robustness of feature modeling, we introduce a variant of the temporal adapter that implements multiple temporal adaptations in parallel while maintaining structural simplicity. Experimental results on real-world datasets demonstrate that the ALL-RP model surpasses state-of-the-art models in predicting both individual user reposting behavior and group sharing behavior, with performance gains of 2.81% and 4.29%, respectively.

KEYWORDS: Reposting prediction; large language model; semantic adaptation; temporal adaptation

1 Introduction

Microblogging platforms such as X and Sina Weibo allow users to share and interact with content online, making reposting a crucial factor in driving information dissemination [1]. Consequently, accurately predicting user reposting behavior holds significant importance across diverse applications. For instance, advertising and marketing enterprises rely on predicting users' reposting intentions to gauge the reach of advertising campaigns [2,3], while online social networking platforms aim to forecast users' propensity to repost when suggesting content to foster community engagement and enhance user retention rates [4].

In user reposting behavior understanding, a common practice is to use Deep Neural Network (DNN) to learn post content and social relationship information in microblogging data to mine users' interests and interaction preferences. For example, Wang et al. [5] combine autoencoders and convolutional neural networks to model user attributes and post semantics to mine users' reposting preferences; Sun et al. [6] devise a pluggable causal attention module for reposting prediction models to understand the causal association between features. However, these methods have at least two drawbacks: (1) Limited by model size



and data scale, the current reposting prediction model cannot fully capture the textual and visual knowledge within microblogging data, showing its relatively general natural language and vision understanding ability, resulting in unsatisfactory prediction performance in various scenarios; (2) Current reposting prediction models generally use random walks or graph neural networks to learn social relationship features between users on sampled online social networks, ignoring that the large scale and complex structure of real online social networks may exceed the representation capabilities of these features.

To overcome the drawbacks, an effective solution is to introduce the Large Language Model (LLM) into the reposting prediction model. Because recent research shows that the rich training corpus and advanced large-scale neural network structure enable LLMs to exhibit emergent ability [7] that traditional deep learning models do not have, and it can well understand multimodal data including natural language, vision, and graph structure information [8]. For example, the LLM introduced in [9] demonstrates optimal performance across various natural language understanding tasks, including question answering, common sense reasoning, and in-context reading comprehension, particularly in a zero-shot learning scenario; Li et al. [10] demonstrated that aligning features between linguistic and visual LLMs enables near state-of-the-art or superior performance across tasks including image-text retrieval, image classification, object detection, and visual grounding; The work in [11] integrates different prompt engineering techniques to demonstrate that the LLM can effectively perform a range of graph reasoning tasks, including node connectivity, topological sorting, shortest path, and maximum flow. However, when using the LLM to solve real-world tasks, current work mainly focuses on learning static information [12,13] or simple temporal pattern [14,15], ignoring user features are temporal and multi-dimensional in the reposting prediction task. Grasping this temporal aspect is essential for the LLM to accurately capture the user's reposting behaviors. We will provide a formal introduction to this issue in [Section 3.1.3](#).

In this paper, we propose a novel approach called Adapt pre-trained Large Language Model for Reposting Prediction (ALL-RP). The ALL-RP model leverages the LLM's capability to represent multimodal user data while enhancing its ability to recognize temporal dynamics. Specifically, we begin by freezing the pre-trained LLM and incorporating a lightweight Adapter [16] after the self-attention layer in each LLM block to enable semantic adaptation. This adjustment allows the LLM to transfer its knowledge of text, vision, and graph data to microblogging contexts. To further enhance the LLM's ability to detect changes in user topic interests and social preferences over time, we introduce a Temporal Adapter (TA) and place it after the feed-forward layer in each LLM block. The ALL-RP model performs a batch-splitting operation on user data, distributing historical repost content and interaction patterns across individual TAs, enabling the model to capture temporal patterns across various dimensions. Additionally, to strengthen the robustness of temporal modeling, we develop a variant of the TA called the Multi-head Temporal Adapter (MTA). MTA segments user features, applies temporal adaptation to each segment independently, and merges the results to form the final representation. Notably, the MTA's structure remains as simple as the Adapter [16]. In summary, this work addresses the following research questions (RQs) to enhance understanding of the proposed method and assess its effectiveness:

- **RQ1:** How can limited data be used to train a repost prediction model that effectively captures the text, image, and graph structure features in online social network data?
- **RQ2:** How can we equip LLMs with the ability to perceive time-varying features across different dimensions of user attributes, such as user topic interests and social preferences?
- **RQ3:** Does the proposed repost prediction model, ALL-RP, outperform state-of-the-art models? What are the specific effects of the semantic adapter, temporal adapter, and multi-head temporal adapter on the repost prediction performance of the ALL-RP model?

The main contributions of this study and answers to the research questions (ARQs) are summarized as follows:

- **ARQ1:** LLMs with large-scale advanced neural networks can have a deep understanding of textual, visual and graph structure information after pre-training on massive data. By inserting a small number of trainable parameters into LLMs, the ALL-RP model proposed in this paper can fine-tune LLMs on limited training data and transfer the powerful feature representation ability of LLMs to the reposting prediction task.
- **ARQ2:** The proposed ALL-RP model introduces a temporal adapter that employs an attention mechanism to capture intrinsic connections between user features over time. Recognizing that the temporal dynamics of different user attributes vary, the model performs batch-splitting on user data, enabling multiple independent adapters to learn temporal patterns specific to each user attribute.
- **ARQ3:** This study compares the repost prediction performance of the ALL-RP model with state-of-the-art models across multiple real-world datasets. Through ablation experiments, the study further evaluates the contribution of each module to the model's performance. Results indicate that the ALL-RP model outperforms the leading baseline by approximately 2.81% in individual repost predictions and 4.29% in group repost predictions, with the semantic adapter, temporal adapter, and multi-head temporal adapter each positively impacting the model's performance.

2 Related Work

2.1 Large Language Model

The Transformer model introduced by Vaswani et al. [17] serves as the foundation for contemporary Large Language Models (LLMs). Its ability to parallelize and scale effectively allows LLMs to handle vast datasets and expand to include billions or even trillions of parameters. Early implementations of Transformer-based LLMs, such as BERT [18] and GPT-1 [19], featured specialized pre-training tasks on extensive corpora. These context-aware word representations, which were pre-trained, have proven to be highly effective as versatile semantic features, significantly enhancing the performance of various natural language processing tasks. These foundational studies have led to a multitude of subsequent developments, including notable models like ChatGPT and LLaMA 3. Recent findings indicate that increasing the scale of LLMs—whether through model size or data volume—can result in unexpected capabilities, referred to as emergent abilities [7], enabling them to tackle complex challenges. For instance, research [20] employs an attention mechanism to integrate feature representations from both low-resource language LLMs and multilingual LLMs, enabling the LLM-based AuthorNet model to gain a nuanced understanding of low-resource language while benefiting from cross-language knowledge transfer. This approach allows the AuthorNet model to effectively address out-of-vocabulary challenges, achieving state-of-the-art performance in low-resource authorship attribution tasks; research [11] demonstrates that LLMs can effectively handle several graph reasoning tasks, such as cycle detection, topological sorting, and bipartite graph matching, indicating their proficiency in interpreting graph structures. Consequently, the integration of LLMs in graph-related applications, such as question answering systems [12] and recommender systems [21], is becoming increasingly prevalent.

2.2 Reposting Prediction Model

Comprehending user reposting behavior is crucial for effective opinion detection and the development of recommender systems [22,23]. Recent approaches predominantly utilize deep neural networks to extract features from microblogging data and to predict user reposting actions, informed by user interests in topics and preferences for social interactions. For instance, study [24] employed a radial basis function network to capture the high-dimensional nonlinear characteristics of social connections and user attributes, thereby identifying reposting inclinations. Another study [25] structured microblogging data into a heterogeneous

reposting graph and introduced an innovative multi-modal neural network model to extract features from this complex graph. This model combined convolutional neural networks and long short-term memory networks to analyze post features while utilizing a user embedding matrix to understand social preferences. Additionally, research [26] presented a multimodal framework that integrates post features, social relationships, and temporal aspects to forecast the timing of reposting actions. Furthermore, study [4] implemented a multi-layer Transformer model as a feature encoder, effectively learning a combined representation of textual and numerical attributes. They also embraced a multi-task learning strategy to accommodate the reposting preferences of diverse user groups, ensuring that the features derived from the encoder accurately reflect the reposting tendencies of users with varying identities.

2.3 Summary

Current reposting prediction methods are limited by their model size and the scale of training data, which hinders their ability to accurately capture users' reposting preferences and, consequently, affects the models' overall performance. Furthermore, existing LLM fine-tuning methods lack the capability to capture time-varying patterns across the multi-dimensional attributes of users in online social networks. To address these limitations, this paper presents the ALL-RP model along with several key innovations. **First**, it integrates both text and image LLMs as backbone models to enhance feature representation in the repost prediction model, enabling strong generalization after fine-tuning. **Second**, it introduces a temporal adapter, allowing the LLMs to perceive temporal patterns in user attributes—such as topic interests and social preferences—that evolve over time. **Third**, to improve robustness in temporal modeling, the model includes a multi-head temporal adapter, enabling it to learn multiple temporal patterns of user attributes concurrently while maintaining a streamlined structure. **Finally**, the predictive capabilities of the ALL-RP model are evaluated against leading models for both individual and group reposting behaviors across multiple real-world datasets, with further exploration into potential applications such as low-resource authorship attribution and recommendation systems.

3 Methodology

3.1 Preliminary

Since the introduction of BERT [18] and GPT-1 [19], transformer-based Large Language Models (LLMs) have found extensive application across a range of data mining tasks [9,11,21]. In this study, we concentrate on harnessing the robust natural language, vision, and graph embedding capabilities of LLMs to enhance reposting prediction tasks. We also compare our approach with leading reposting prediction models to evaluate its effectiveness.

3.1.1 Multimodal Inputs

More specifically, this work predicts user u 's reposting behavior for query post p_q based on user u 's past reposting content and interaction relationships. In order to facilitate the LLM to process the input data, we represent the input uniformly as a text or image sequence. The user u 's past multimodal reposting content are organized as a text and image sequence P , $P = \{[t_0^u; \dots; t_W^u], [v_0^u; \dots; v_W^u]\}$, where t_w^u and v_w^u is the text and image in the reposting record of user u , $0 \leq w \leq W$. Meanwhile, the reposting interactions of user u are represented as $[r_0^u; \dots; r_I^u]$. Inspired by literature [11,27], we describe users' reposting interactions r_i^u at different times as natural language in the following format: " u reposted a post from a_p on d_p ", $0 \leq i \leq I$. Here, a_p is the author of post p , d_p is the reposting date.

3.1.2 LLM Block

Large Language Models (LLMs) process text as a sequence of tokens. For an input text $x \in \mathbb{R}^N$, the LLM first tokenizes it into T tokens and maps each token to a D -dimensional embedding through a trainable embedding layer. Here, N denotes the length of the text sequence, while the text embeddings are represented as $x_t \in \mathbb{R}^{T \times D}$. It is important to note that N typically differs from T because LLMs often insert special tokens, such as [SEP] and [PAD], into the tokenized sequence. These tokens aid the model in interpreting the structure and organization of the input text. For an input image $x \in \mathbb{R}^{H \times W \times C}$, the LLM segments the image into T non-overlapping patches and embeds them into D -dimensional vectors. In this scenario, $(H \times W)$ represents the image resolution, and C indicates the number of channels. The image embedding is denoted as $x_v \in \mathbb{R}^{T \times D}$, where $T = HW/P^2$ and P is the patch size. To incorporate positional information for both text tokens and image patches, positional embeddings $E_{pos} \in \mathbb{R}^{T \times D}$ are added to x_t and x_v , resulting in $h_0^t = x_t + E_{pos}$ and $h_0^v = x_v + E_{pos}$. This combined representation h_0 is then processed through a series of LLM blocks. Each block consists of multi-head attention (MHA) and feed-forward layers (FFL), complemented by layer normalization (LN) and skip connections, as illustrated in Fig. 1a. The computation of a standard LLM block can be expressed as follows:

$$h'_l = h_{l-1} + \text{MHA}(\text{LN}(h_{l-1})), \quad (1)$$

$$h_l = h'_l + \text{FFL}(\text{LN}(h'_l)), \quad (2)$$

where h_{l-1} and h_l represent the input and output of the l -th LLM block, respectively. Ultimately, the representation obtained from the final LLM block serves as the input for the classification head, which is responsible for generating the predictions.

3.1.3 Challenges in Temporal User Features Learning

User features are embedded in the content and interactions of posts they have shared over time. Therefore, understanding reposting behavior necessitates the model to develop effective semantic representations at each point in time (semantic modeling) and to infer the temporal structure of information across different time periods (temporal modeling). A crucial aspect of leveraging an LLM for reposting prediction tasks is the implementation of temporal modeling. Most recent studies [12,13] have focused on semantic-only models, which utilize LLMs to process user data independently at various times. For an input representation $x \in \mathbb{R}^{S \times T \times D}$, where S represents the number of time slices, these semantic-only models generate S semantic features, each corresponding to a specific time slice. These features are then averaged to model temporal dynamics for final predictions. To improve temporal modeling capabilities, some approaches [14,15] have introduced semantic-time models by incorporating additional temporal modules into LLMs. However, these methods are limited in that they can only learn temporal patterns within a single dimension of user attributes. In this work, we investigate how to effectively adapt LLMs for the reposting prediction task, enhancing their ability to capture temporal user features, including topic interests and social preferences.

3.2 Semantic Adaptation

Given that LLMs have been trained on extensive datasets and have exhibited strong transferability to various downstream tasks, they can effectively achieve robust semantic modeling for reposting prediction with minimal fine-tuning. Drawing inspiration from efficient fine-tuning techniques in natural language processing [28], we utilize the Adapter method [16] due to its straightforward implementation. As illustrated in Fig. 2a, the Adapter employs a bottleneck architecture, consisting of two linear layers with an activation function between them. The initial linear layer reduces the input dimensionality, while the second linear layer reconstructs it back to the original dimension.

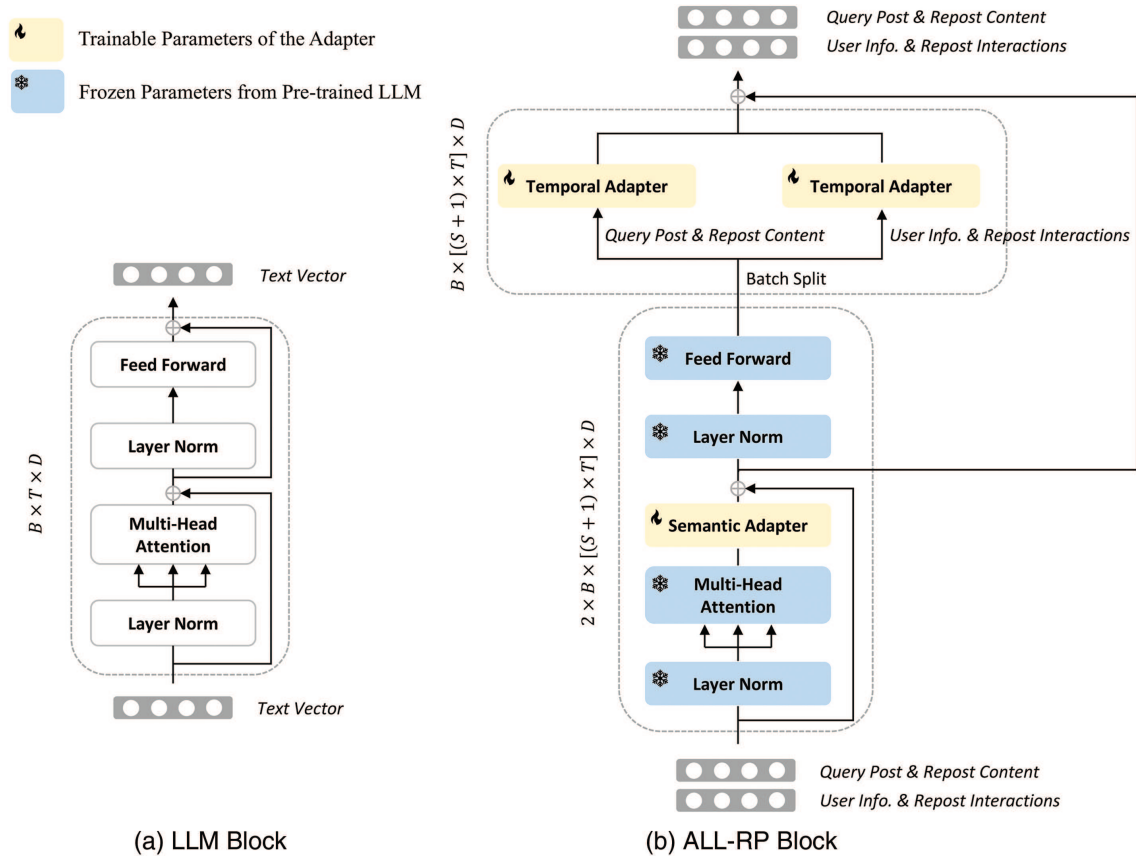


Figure 1: The structure of the LLM and the proposed ALL-RP model

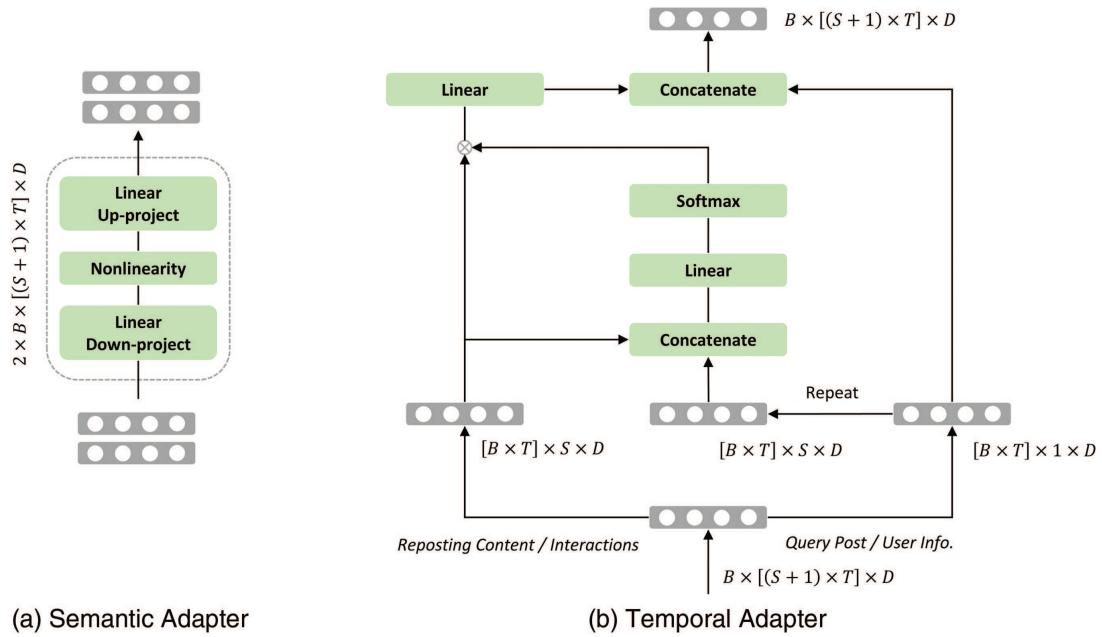


Figure 2: The structure of the semantic and temporal adapter

For a query post p_q , when predict a user u 's reposting behavior, we first represent the information about the user u as two sequences, i.e., $\mathcal{I} = [p_q; p_1^u; \dots; p_S^u]$ and $\mathcal{S} = [r_{au}; r_1^u; \dots; r_S^u]$, where r_{au} is a description of the user and query post publisher information: "the user is u , the query post author is a ". From \mathcal{I} and \mathcal{S} , LLM can perceive users' topic interests and social preferences. Then, under the processing of LLM's tokenizer and embedding layer, the input text sequences are converted into embedding $x_u \in \mathbb{R}^{2 \times [(S+1) \times T] \times D}$. If we input B query posts at a time while fine-tuning, then $x_u \in \mathbb{R}^{2 \times B \times [(S+1) \times T] \times D}$. Finally, to tailor the pre-trained semantic features to the specific characteristics of microblogging data, we incorporate an Adapter after the multi-head attention layer, as depicted in Fig. 1b. We refer to this process as semantic adaptation. During the training phase, all other layers of the LLM remain frozen, with only the adapters being updated.

3.3 Temporal Adaptation

User features can evolve over time, making it vital to capture these time-varying patterns and explore their correlations with query posts to accurately predict user reposting behavior. Recent approaches [14,15] often enhance pre-trained LLMs by integrating new temporal modules to effectively capture temporal information. However, these methods typically focus on learning temporal patterns within a single dimension, whereas user features encompass multiple dimensions, such as topic interest and social preference. The challenge lies in modeling the multi-dimensional temporal patterns of user features, which is crucial for fine-tuning LLMs for the reposting prediction task.

To tackle this issue, we propose a novel model called ALL-RP. More specifically, we perform a batch split operation on the embedding after it passes through the feed forward layer, and send the embedding containing the user's topic interests and social preference information to different temporal adapters to learn their temporal patterns respectively, see Fig. 1b. This facilitates the proposed ALL-RP model to learn multi-dimensional temporal features of user attributes. Importantly, we make a new design for the temporal adapter, see Fig. 2b. Now given the embedding $x_{qr} \in \mathbb{R}^{B \times [(S+1) \times T] \times D}$ of the query post and user past reposting content, we first separate and reshape them to $x_q \in \mathbb{R}^{[B \times T] \times 1 \times D}$ and $x_r \in \mathbb{R}^{[B \times T] \times S \times D}$, where B represents the batch size, T denotes the number of post tokens, and S indicates the number of reposts. We repeat x_q to $x'_q \in \mathbb{R}^{[B \times T] \times S \times D}$ and concatenate it with x_r to $x'_{qr} \in \mathbb{R}^{[B \times T] \times S \times 2D}$. We then input x'_{qr} into a linear layer, which aims to learn the relationship between the query post and the posts forwarded by users across S different time points. The generated temporal pattern matrix is $p \in \mathbb{R}^{[B \times T] \times S \times S}$. Finally, we perform temporal modeling on reposting sequence x_r . These processes can be formulated as

$$p = \text{Softmax}(x'_{qr} W_t), \quad (3)$$

$$x'_r = p x_r W_p, \quad (4)$$

where $W_t \in \mathbb{R}^{2D \times S}$ and $W_p \in \mathbb{R}^{D \times D}$ are the parameters of linear layers. The output of temporal adapter is the concatenation of x_q and the learned x'_r . Similarly, the ALL-RP model uses the temporal adapter to learn the time pattern of user social preferences x'_s .

Multi-head temporal adapter. To enhance the robustness of the proposed ALL-RP model, we partition the features into H segments, perform the temporal adaptation separately, and combine the results to obtain the final representation:

$$p^{(h)} = \text{Softmax}(x'_{qr} W_t^{(h)}), \quad (5)$$

$$x'_r = [p^{(0)} x_r^{(0)}, \dots, p^{(H-1)} x_r^{(H-1)}] W_p, \quad (6)$$

where $0 \leq h \leq H$, $W_t^{(h)} \in \mathbb{R}^{2D \times S}$, $x_r^{(h)}$ is the features in the h -th segment, $[\cdot; \cdot]$ is the concatenation operation. Note that, when implementing the proposed ALL-RP model, we can replace multiple $W_t^{(h)}$ with a linear layer of parameter $W_t^H \in \mathbb{R}^{2D \times HS}$, so the multi-head temporal adapter (MTA) still contains only two linear layers and a nonlinear activation function. This ensures that our MTA has the same simple structure as the original Adapter [16] while efficiently learning the temporal patterns of the sequence.

3.4 ALL-RP Incorporating Text and Visual Content

With the rise of online social networks, users increasingly rely on multimedia to express themselves, making text and images integral components of post content. To improve the ALL-RP model's ability to handle visual information, especially on platforms rich in visual content like Flickr and REDnote, the model incorporates both text and visual backbone architectures (see Fig. 3). In this framework, user social relationships are represented as natural language (as described in Section 3.1.1) and fed into text backbone models, such as GPT [19] or LLaMA, alongside the post's text content. Meanwhile, image data from the post is processed by the visual backbone models, such as ViT [29] or SegFormer [30].

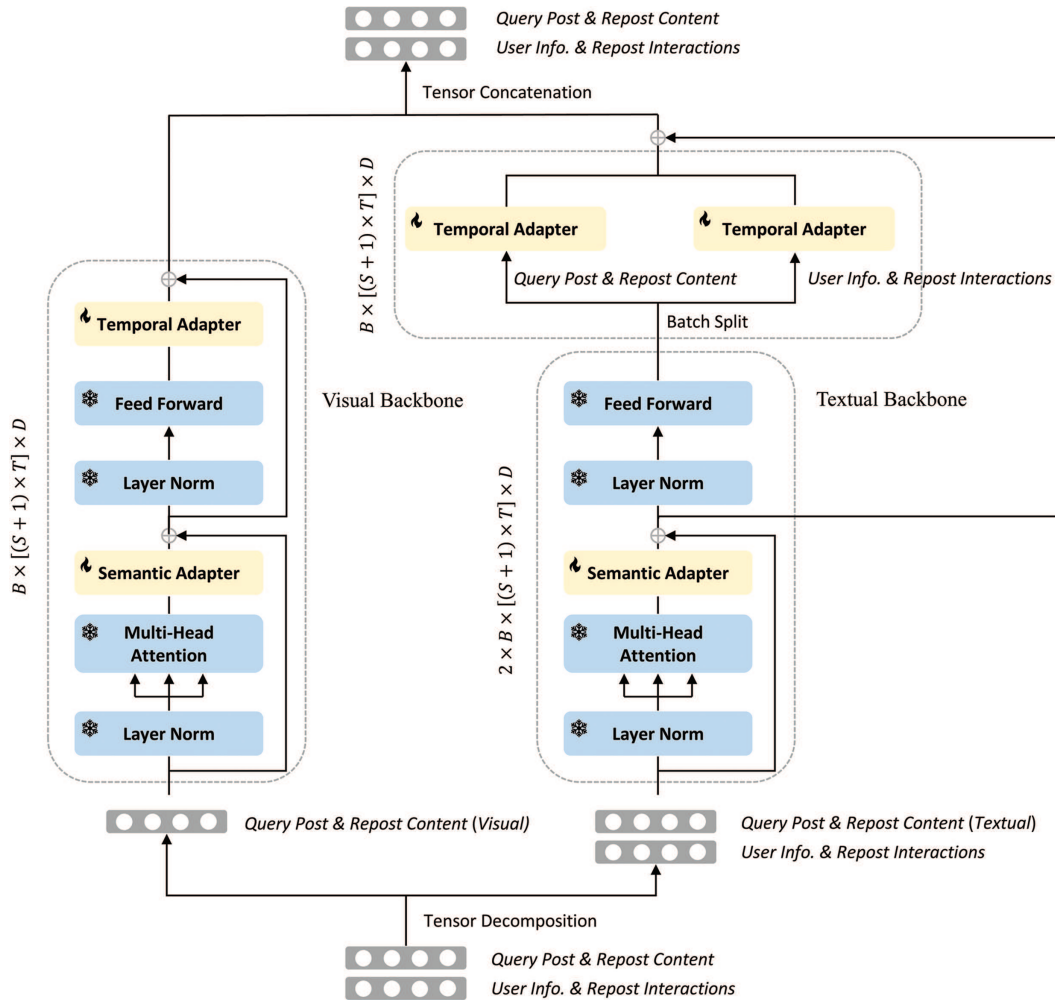


Figure 3: ALL-RP model with textual and visual backbones

3.5 Reposting Prediction

Each block of the proposed ALL-RP model will do the semantic and temporal adaptations, and the output $x_{rp} \in \mathbb{R}^{2 \times B \times [(S+1) \times T] \times D}$ of the last block will serve as the basis for reposting prediction. The classifier module in our ALL-RP model is composed of two linear layers, with a ReLU activation function applied between them. Formally, for an input x_{rp} , the output \hat{y} is computed as follows:

$$h_{rp} = \text{ReLU}(x_{rp} W_r^{(1)} + b_r^{(1)}), \quad (7)$$

$$\hat{y} = h_{rp} W_r^{(2)} + b_r^{(2)}, \quad (8)$$

where $W_r^{(1)}$ and $W_r^{(2)}$ are the weight matrices, and $b_r^{(1)}$ and $b_r^{(2)}$ are the bias vectors.

When predicting the reposting behavior of a single user, the model is framed as a binary classification model. The model is fine-tuned using the cross-entropy loss, which is defined as follows:

$$\mathcal{L}_c = -\frac{1}{B} \sum_{i=1}^B y_i \log(\hat{y}_i), \quad (9)$$

where y_i is the user reposting behavior, \hat{y}_i is the predicted probability for the user behavior, and B is the batch size. When predicting the spread of a post on a microblogging service—specifically, the number of times it is reposted by the user group—the model is framed as a regression task and fine-tuned using Huber loss:

$$\mathcal{L}_r = \frac{1}{B} \sum_{i=1}^B l_i. \quad (10)$$

$$\text{If } |y_i - \hat{y}_i| < \delta, l_i = 0.5 (y_i - \hat{y}_i)^2; \text{ otherwise, } l_i = \delta (|y_i - \hat{y}_i| - 0.5\delta).$$

4 Experimental Setups

4.1 Datasets

In order to comprehensively assess the effectiveness of the proposed ALL-RP model, we perform experiments using the real-world datasets sourced from X and Flickr.

Archive Team Twitter Grabs (ATTG)¹. X, formerly known as Twitter, is one of the world's most popular microblogging platforms. The ATTG dataset comprises archived Twitter streams from September 2011 to January 2023. For this experiment, we use data from August 2022 to November 2022, which includes information on approximately 69,053,717 X users. To ensure robust experimental results, we generate three samples from this dataset using the following procedure. First, we randomly select 10,000 users as our prediction targets. Second, we identify reposts $\mathcal{R}_t = \{\langle p_1, d_1, u_1, a_1 \rangle, \dots, \langle p_n, d_n, u_n, a_n \rangle\}$ as positive instances, where each post $\langle p_i, d_i, u_i, a_i \rangle$, $1 \leq i \leq n$, represents the content p_i , date d_i , reposter u_i , and original poster a_i . When training the reposting prediction model to forecast u_i 's behavior at d_i , only posts prior to d_i are used as u_i 's repost history, preventing information leakage. Third, we randomly select posts by a_i that were not reposted by u_i as negative instances, maintaining a 1:1 ratio between positive and negative instances. For each sample, the training set covers a span of three months, while the testing set includes data from the 10 days immediately following the training period (see Table 1). Additionally, repost records are used to extract temporal repost interactions $\mathcal{I}_t = \{\langle d_1, u_1, a_1 \rangle, \dots, \langle d_n, u_n, a_n \rangle\}$ between users. These interactions form the dynamic-interaction graph, with \mathcal{I}_t limited to training set interactions to avoid data leakage. A detailed summary of the experimental datasets is provided in Table 2.

¹<https://archive.org/details/twitterstream>, accessed on 01 October 2024.

Table 1: Division of training and testing sets

	Sample	Training set	Testing set
ATTG	1	08/01/2022-10/31/2022	11/01/2022-11/10/2022
	2	08/06/2022-11/05/2022	11/06/2022-11/15/2022
	3	08/11/2022-11/10/2022	11/11/2022-11/20/2022
TPIC2017	1	03/03/2012-05/31/2012	06/01/2012-06/10/2012
	2	03/08/2012-06/05/2012	06/06/2012-06/15/2012
	3	03/13/2012-06/10/2012	06/11/2012-06/20/2012

Table 2: Overview of the ATTG datasets

Sample	Users	Posters	Pos.	Neg.	Histories
1	3653	7003	9346	9280	11,221
2	3763	7458	10,084	10,007	12,133
3	3910	7567	10,274	10,182	12,790

Note: It is important to note that the number of users in each dataset sample is slightly below 10,000, as some randomly selected users may not have exhibited repost behavior during the dataset collection period.

Temporal Popularity Image Collection (TPIC2017)². TPIC2017 is a large-scale Flickr dataset consisting of 680,000 posts with images from anonymized users, accompanied by three years of photo-sharing history. This dataset encompasses images, user profiles, and associated metadata. In the experiment, we first download the images in the posts through the Flickr API according to the post URLs provided in TPIC2017. Then, the experiment uses the post $p_u = \langle i_u, d_u, m_u \rangle$ shared by user u as the instance in the dataset, where i_u is the image in the post, d_u is the date the post was published, and m_u is the metadata of the post (including the number of times the post has been commented on, the length of the post title, the length of the post description, etc.). Next, we construct the records \mathcal{R}_u that can be used for training when user u publishes post p_u . These records contain all posts published by user u before date d_u . The training set contains the user's posting records within 3 months, and the testing set is the data 10 days after the training set (see Table 1). Finally, the reach of each post within the social network is evaluated by the number of times it is viewed by other users:

$$s = \log \frac{r}{d} + 1. \quad (11)$$

Here, s represents the normalized value, r denotes the photo's view count, and d indicates the number of days since the photo was uploaded. The statistical information of the experimental dataset is shown in Table 3.

²<https://github.com/social-media-prediction/TPIC2017>, accessed on 01 October 2024.

Table 3: Overview of the TPIC2017 datasets

Sample	Users	Instances	Reach of sharing	Histories
1	288	15,763	[0, 16]	66,185
2	288	15,543	[0, 16]	66,058
3	287	15,228	[0, 16]	65,689

4.2 Baselines

To evaluate the effectiveness of the ALL-RP model, we compare it against the following reposting prediction baselines in our experiments:

Random: This model generates predictions randomly. For predicting whether a user will repost a specific query post, it randomly chooses between ‘repost’ and ‘do not repost.’ When estimating the reach of sharing within a user group, it randomly selects a floating-point number between 0 and 16 as the prediction result. Any model capable of learning user reposting behavior patterns from microblogging data should outperform the random model.

Linear Regression (LR): LR is a classic machine learning model. In this experiment, pre-trained BERT and ViT [29] models are used as the text and visual content embedding layers for LR. The model uses features from the query post and the average pooled features of the user’s past reposted content as the basis for prediction.

Neural network-based: Models such as AMNL [25], DFMM [26], and GFCI [31] leverage neural networks to learn features from multimodal data for prediction tasks. For example, the GFCI model initially employs a pre-trained BERT model and a convolutional neural network to learn user interest representations from their history. It then uses an attention mechanism and a dynamic graph neural network to explore interest similarity and social relationships between users, capturing their reposting preferences.

LLM-based: Models including TALLRec [32] and A-LLMRec [33] map users and posts to the feature space of LLMs through fine-tuning and alignment. These approaches leverage the LLM’s powerful ability to understand multimodal information, enhancing the perception of users’ reposting patterns.

4.3 Model Configuration

The baseline LR model is implemented using Scikit-learn with default parameters and training settings. As the focus of our work is not on the exact timing of reposting behavior, we exclude the time encoding module from the DFMM [26] model. To account for differences between datasets, modifications are made to the AMNL [25], GFCI [31], TALLRec [32], and A-LLMRec [33] models. For the ATTG dataset, which lacks visual content, the AMNL model omits the CNN module responsible for processing images, while the TALLRec and A-LLMRec models utilize GPT-Neo³ as the backbone model. Conversely, for the TPIC2017 dataset, which lacks social relationship information but includes images, the AMNL and GFCI models remove their respective user embedding matrices and dynamic graph neural networks. In these cases, the TALLRec and A-LLMRec models incorporate both text and visual backbone models (ViT⁴ and GPT-Neo). For consistency, the proposed ALL-RP model uses the same backbone configuration as TALLRec and A-LLMRec across different datasets. All baseline models and the ALL-RP model are implemented in PyTorch. AdamW [34] is used as the optimizer, with the learning rate selected from $\{10^{-2}, 10^{-3}, 10^{-4}, 10^{-5}\}$. The evaluation metrics include Accuracy (A), F1-score (F), Mean Absolute Percentage Error (MAPE), and Mean

³<https://huggingface.co/EleutherAI/gpt-neo-1.3B>, accessed on 01 October 2024.

⁴<https://huggingface.co/google/vit-large-patch16-224-in21k>, accessed on 01 October 2024.

Relative Square Error (MRSE). For non-LLM baselines, the batch size is set to 64, with early stopping applied when the F1-score or MAPE does not improve for 10 consecutive epochs. For LLM-based baselines and the ALL-RP model, training runs for 30 epochs with a batch size of 12. The learning rate is initially warmed up over the first three epochs before following a cosine decay schedule. In the ALL-RP model, the hidden layer dimension of the semantic adapter is set to one-quarter of the input dimension, and the temporal adapter uses 8 heads. The experiments are conducted on an Ubuntu 20.04 platform with 90 GB RAM, using a single NVIDIA V100 GPU for training.

5 Results

5.1 Prediction Comparison

To demonstrate the effectiveness of the proposed ALL-RP model, as outlined in [Section 3](#), we compare its performance against state-of-the-art baselines. The prediction performances of the ALL-RP model and the baselines are presented in [Tables 4](#) and [5](#). From the results, we can find that the LR model demonstrates superior performance compared to the Random baseline. This observation indicates that LLMs can extract useful knowledge for the LR model from microblogging data through their powerful feature representation abilities, even without fine-tuning or alignment. However, because the LR model can only capture shallow linear relationships between features, its performance is inferior to neural network-based models in predicting both micro-level individual user reposting behavior (see [Table 4](#)) and macro-level user group sharing behavior (see [Table 5](#)). For example, the AMNL model outperforms the LR model by approximately 6.23% in terms of Accuracy and 15.18% in MAPE.

Table 4: Prediction performances on ATTG datasets

Model	Sample 1		Sample 2		Sample 3	
	A	F	A	F	A	F
Random	0.511	0.513	0.509	0.512	0.515	0.517
LR	0.589	0.575	0.611	0.608	0.605	0.605
AMNL	0.639	0.621	0.637	0.635	0.641	0.628
DFMF	0.628	0.616	0.612	0.626	0.625	0.594
GFCI	0.641	0.626	0.650	0.641	0.643	0.646
TALLRec	0.654	0.652	0.680	0.678	0.670	0.642
A-LLMRec	0.665	0.676	0.669	0.657	0.661	0.636
ALL-RP	0.677	0.691	0.693	0.709	0.681	0.662

Table 5: Prediction performances on TPIC2017 datasets

Model	Sample 1		Sample 2		Sample 3	
	MAPE	MRSE	MAPE	MRSE	MAPE	MRSE
Random	>10 ⁴	>10 ⁴	>10 ⁴	>10 ⁴	>10 ⁴	>10 ⁴

(Continued)

Table 5 (continued)

Model	Sample 1		Sample 2		Sample 3	
	MAPE	MRSE	MAPE	MRSE	MAPE	MRSE
LR	0.368	0.272	0.427	0.300	0.372	0.266
AMNL	0.316	0.224	0.348	0.174	0.324	0.197
DFMF	0.327	0.228	0.322	0.187	0.353	0.215
GFCI	0.296	0.186	0.319	0.185	0.313	0.192
TALLRec	0.287	0.169	0.297	0.166	0.280	0.180
A-LLMRec	0.271	0.165	0.292	0.170	0.285	0.178
ALL-RP	0.269	0.145	0.276	0.152	0.266	0.147

For LLM-based models, they demonstrate notable improvements over traditional neural network-based models across multiple datasets. For example, when predicting individual user reposting behavior (see Table 4), the TALLRec model achieves an accuracy approximately 3.61% higher than the GFCI model. Similarly, for user group predictions (see Table 5), the TALLRec model improves the MAPE index by around 6.83% compared to the GFCI model. These gains can be attributed to the extensive pre-training data and the advanced large-scale neural network architecture, which provides the LLM model with enhanced feature representation capabilities. After fine-tuning and alignment, LLM-based reposting prediction models exhibit a superior understanding of microblogging data. Among the LLM-based models, the performance of the A-LLMRec model is comparable to that of TALLRec, yet both fall short of the ALL-RP model proposed in this paper. Specifically, the ALL-RP model outperforms the A-LLMRec model by 2.81% in predicting individual reposting behavior and by 4.29% in predicting group reposting behavior. This demonstrates that, beyond fine-tuning LLMs, modeling users' multi-dimensional time-varying features can further unlock the potential of LLMs for reposting behavior prediction tasks.

5.2 Ablation Study

As discussed in Section 3, the proposed ALL-RP model introduces a semantic adapter (SA) and a temporal adapter (TA) to capture post semantics and dynamic user features for the reposting prediction task. To further improve temporal modeling robustness, we propose a multi-head temporal adapter (MTA). To evaluate the contributions of these components to model performance, we conduct an ablation study. First, we train the classification module of the ALL-RP model without adapting the features learned by the backbone model; this variant is labeled ALL-RP (w/o SA & TA). Second, we perform only semantic adaptation on the features captured by the backbone model, marking this variant as ALL-RP (w/o TA). Third, we insert only the MTA into the backbone model, labeling this as ALL-RP (w/o SA). Finally, we create a variant where the ALL-RP model replaces the MTA with the TA, as described in Eqs. (3) and (4), enabling both semantic and temporal adaptations but without parallel modeling. This variant is denoted as ALL-RP (w/o MTA). The results are presented in Fig. 4.

From the results, it becomes evident that ALL-RP (w/o SA & TA) exhibits the poorest prediction performance regardless of the datasets. As shown in Fig. 4a, when predicting the reposting behavior of individual users for query posts, it performs 8.09% lower than ALL-RP and 3.13% lower than ALL-RP (w/o TA) which does not take into account the fact that users' topic interests and social preferences may change over time. A similar situation also occurs when predicting the behavior of user groups, see Fig. 4b.

This indicates that when applying the pre-trained large language model to the reposting prediction task, the parameter fine-tuning is required in order to identify microblogging user-related features. Meanwhile, by comparing the performance of ALL-RP (w/o TA) and ALL-RP, it is clear that learning information about the time dimension of user features is essential for ALL-RP model. Without temporal adaptation, the performance of the ALL-RP model on the ATTG and TPIC2017 datasets decreases by 5.12% and 10.82%, respectively. Furthermore, upon removing the semantic adapter, the Accuracy and MAPE of ALL-RP model demonstrate varying degrees of reduction across different cases. On average, the ALL-RP (w/o SA) model exhibits 2.48% lower Accuracy and 4.79% lower MAPE compared to ALL-RP model. These findings emphasize the significance of the semantic adapter in assisting ALL-RP model in better perceiving the semantic information of text and images. Lastly, due to take into account the semantic differences between the backbone model training corpus and the microblogging data as well as the temporal properties of user features, the ALL-RP (w/o MTA) model outperforms all other variants. However, limited by the robustness of the temporal modeling module, the performance of the ALL-RP (w/o MTA) model on the ATTG and TPIC2017 datasets is still about 1.70% and 3.83% lower than that of the ALL-RP model, respectively. Based on the above results, we can know that the semantic adapter as well as multi-head temporal adapter are all integral parts of the ALL-RP model.

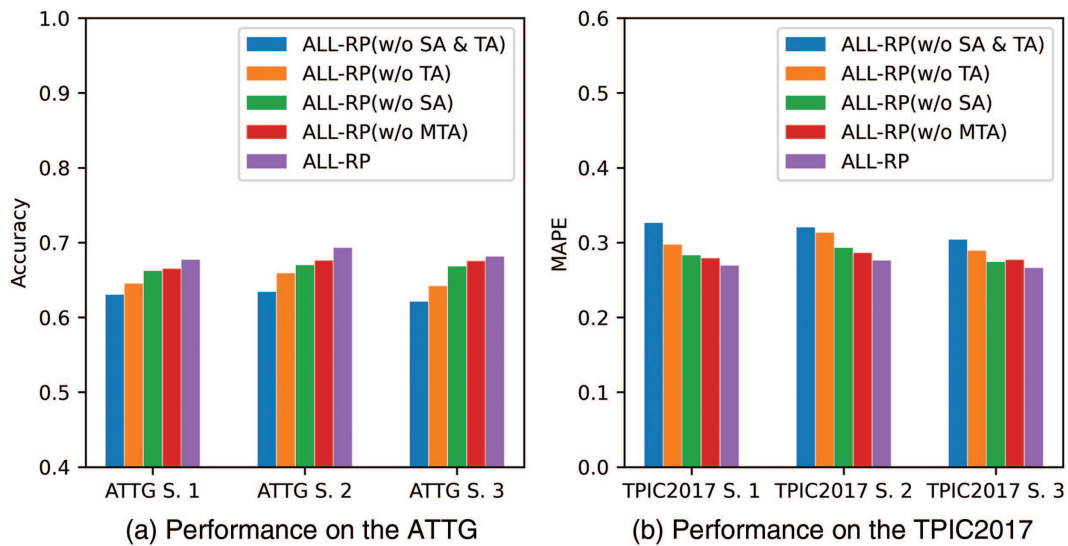


Figure 4: Prediction performance of different variants of ALL-RP model

5.3 Model Analysis

Different Learning Rates. By adjusting the learning rate of the ALL-RP model, we can effectively control the speed of tunable parameter updates. In this section, we explore how the learning rate impacts the prediction performance of the ALL-RP model. Table 6 presents the Accuracy and MAPE metrics for the first sample of the ATTG and TPIC2017 datasets, respectively. The results show that a learning rate of 1×10^{-4} yields the best performance for training the ALL-RP model. Both excessively high and low learning rates negatively affect performance, a pattern consistent with other LLM-based models [32,33]. A high learning rate causes the model to quickly overfit the training data, leading to poor performance on test data. On the other hand, a very low learning rate requires numerous iterations to improve performance and may get stuck in local optima of the loss function, resulting in subpar final outcomes.

Table 6: Effect of learning rate

Learning rate	Accuracy	MAPE
1×10^{-2}	0.646	0.321
1×10^{-3}	0.669	0.288
1×10^{-4}	0.677	0.269
1×10^{-5}	0.653	0.275

Data Efficiency. We also demonstrate the effectiveness of the ALL-RP model with varying amounts of training data using the first sample dataset of ATTG. For fairness, both the ALL-RP and A-LLMRec models utilize GPT-Neo as the backbone model. As shown in Table 7, the ALL-RP model consistently outperforms the state-of-the-art A-LLMRec model, even when training data is limited. This indicates that the proposed ALL-RP model is more data-efficient and less susceptible to overfitting. Notably, the advantage of the ALL-RP model becomes more pronounced as the amount of training data decreases. For instance, when the training data is reduced to 10% of the original amount, the performance improvement of the ALL-RP model over the A-LLMRec model increases from 1.80% to 4.79%.

Table 7: Data efficiency comparison

Training data	ALL-RP	A-LLMRec
10%	0.613	0.585
20%	0.649	0.630
50%	0.668	0.652
100%	0.677	0.665

Position of Adapters. By default, adaptations are applied in every LLM block. In this study, we investigate the impact of adding semantic and temporal adapters at different layers of the model. Using the pre-trained GPT-Neo as the backbone (with 24 blocks), we add adapters to three different configurations: the top 12 blocks (near the output), one every two blocks, and the bottom 12 blocks (near the input). As shown in Table 8, applying adapters to the top 12 blocks significantly degrades performance on the ATTG datasets. We hypothesize that this is because shallow layers learn generic representations that strongly influence the deeper layers, making feature adaptation in the bottom layers more crucial. Adding adapters to the bottom 12 blocks achieves performance comparable to applying them in all blocks while reducing the number of parameters by half. This approach may be preferable when training resources are limited.

Error Analysis. The error analysis in this study offers a detailed evaluation of the strengths and limitations of the proposed ALL-RP model compared to the baseline models. Table 9 presents the error values for the ATTG dataset across various methods, including Random, Linear Regression, Neural Network-based models, and LLM-based models. Compared to the random model, the linear regression model significantly reduces the error rate by learning the correlation between query posts and user topic interests. However, its inability to capture complex, non-linear feature relationships and to leverage social information between users results in a higher error rate compared to neural network-based models (e.g., AMNL, DFME, and GFCI). With the advent of LLMs, reposting prediction models such as TALLRec, A-LLMRec, and the

proposed ALL-RP, which utilize LLMs as backbone models, demonstrate superior feature representation capabilities, leading to further reductions in error rates. Notably, the ALL-RP model achieves the lowest error rate by accounting for the time-varying nature of user attributes, such as topic interests and social preferences, through careful modeling of these dynamic characteristics.

Table 8: Effect of position of adapters

Position	Sample 1	Sample 2	Sample 3
Top 12	0.638	0.642	0.644
Skip	0.651	0.667	0.658
Bottom 12	0.669	0.679	0.664
All	0.677	0.693	0.681

Table 9: Error analysis of the ATTG dataset for the proposed ALL-RP model and baselines

Model	Error (%)		
	Sample 1	Sample 2	Sample 3
Random	48.9	49.1	48.5
LR	41.1	38.9	39.5
AMNL	36.1	36.3	35.9
DFMF	37.2	38.8	37.5
GFCI	35.9	35.0	35.7
TALLRec	34.6	32.0	33.0
A-LLMRec	33.5	33.1	33.9
ALL-RP	32.3	30.7	31.9

While the proposed ALL-RP model achieves the lowest error rate, it is not able to accurately predict all user reposting behaviors. To illustrate the challenges, [Table 10](#) presents two query post cases that highlight factors influencing the ALL-RP model's predictions. For Case 1, the ALL-RP model correctly predicts the user's behavior, as the post is written in English—the dominant language on the X platform—and its semantics are effectively captured by the model's backbone, GPT-Neo. However, the model fails to correctly predict Case 2, which contains both Thai and English. Since the backbone model's pretraining corpus primarily consists of English text, the ALL-RP model struggles to comprehend the semantics of both languages and their interactions. Query posts similar to Case 2 pose challenges for the ALL-RP model, often resulting in prediction errors.

Table 10: The query posts of two cases

Case	Content
1	(CN 30th) No money means No animation
2	Food support ที่รอคอยมาแล้วจ้า ติดต่อมิวไทยได้ตามรายละเอียดจ้า. @MSuppasit

Beyond Repost Prediction. The reposting prediction task is closely connected to the low-resource authorship attribution task, as both require a deep analysis of text to capture users' topic preferences, writing logic, and word usage patterns. The LLM-based reposting prediction model, ALL-RP, proposed in this paper, can leverage trainable adapters to transfer the rich knowledge of low-resource languages embedded in LLMs to the authorship attribution task, thereby achieving the desired performance. Furthermore, the reposting prediction task is closely related to the content recommendation task, as both require analyzing user attributes and extracting user interests from past interactions to predict reposting behavior on a query post or purchasing behavior for a product. Crucially, both tasks emphasize the use of temporal information from user interaction histories to track changes in user identity and interests over time, enabling the model to more accurately capture user preferences. To explore additional application scenarios, this section evaluates the performance of the proposed ALL-RP model on the content recommendation task. The experiment frames the recommendation task as a binary classification problem, conducted on Amazon data⁵, and compares the ALL-RP model with state-of-the-art recommendation models (i.e., TALLRec [32] and A-LLMRec [33]), as shown in Table 11. The results demonstrate that the ALL-RP model performs well on the content recommendation task, second only to the A-LLMRec model. The superior performance of A-LLMRec can be attributed to its integration of the powerful feature representation capabilities of LLMs with a pre-trained collaborative filtering recommender, allowing it to obtain high-quality user and product features, thus better capturing user preferences and product similarities.

Table 11: Recommendation performances (Hit@1) on different datasets

Position	ALL-RP	TALLRec	A-LLMRec
Movies and TV	0.351	0.235	0.624
Video games	0.429	0.440	0.528

5.4 Discussion

This study introduces a novel LLM-based reposting prediction model, ALL-RP. The model fine-tunes the pre-trained LLM by integrating a trainable Adapter [16], leveraging its extensive knowledge to address the reposting behavior prediction task. This enables the model to better interpret text, images, and social relationships within microblogging data, enhancing its ability to capture user reposting patterns. A key advantage of the ALL-RP model is its ability to leverage the Temporal Adapter (TA) proposed in this paper to model the time-varying features of users' multidimensional attributes, such as topic interests and social relationships, resulting in superior reposting prediction performance compared to other LLM-based models (see Section 5.1). To further strengthen the model's temporal attribute modeling capabilities, this paper introduces a Multi-head Temporal Adapter (MTA). This module efficiently learns diverse temporal patterns

⁵<https://nijianmo.github.io/amazon/index.html>, accessed on 01 October 2024.

of user attributes in parallel while maintaining a structure as simple as the original Adapter, as demonstrated in the ablation study (see Section 5.2). Moreover, beyond its primary role in reposting prediction, ALL-RP also exhibits strong performance in recommendation tasks (see Section 5.3), indicating its potential applicability for recommendation services on online social platforms.

While ALL-RP demonstrates impressive results, it also has some limitations. Similar to the Adapter model, this work introduces a small number of trainable parameters on top of the pre-trained LLM parameters and fine-tunes the LLM by training these additional parameters for the reposting behavior prediction task. However, to capture the different temporal patterns of user interests and social relationships, this study applies independent temporal adaptations to this information (see Fig. 1). While the proposed TA and MTA modules are as simple as the Adapter structure, their greater number results in the ALL-RP model requiring more parameters to be trained compared to the original Adapter model [16]. This may restrict the use of the ALL-RP model in scenarios with limited computational resources.

6 Conclusion

Forecasting users' reposting behaviors is crucial for optimizing advertising campaigns and enhancing recommender systems. Existing methods generally use deep neural networks to learn and predict user reposting behavior. However, limited by the scale of model and training data, it is difficult for these methods to well perceive user topic and social preference features related to the reposting behavior. To address this problem, we introduce the pre-trained LLM for the reposting prediction task. The extensive parameter count in the LLM allows it to assimilate a wide array of linguistic and factual knowledge from a large-scale corpus. Meanwhile, the proposed ALL-RP model accounts for the semantic specificity of microblogging data and the temporality of user attributes by performing both semantic and temporal adaptation. Notably, when modeling users' complex temporal attributes, the ALL-RP model concurrently learns temporal patterns across multiple dimensions, including topic interests and social preferences, while ensuring model robustness and simplicity. On the public datasets, the proposed ALL-RP model surpasses state-of-the-art models in predicting both individual user reposting behavior and group sharing behavior, with improvements of 2.81% and 4.29%, respectively. In the future, we plan to develop a more lightweight fine-tuning approach for LLMs to improve the training efficiency of the ALL-RP model, thereby enhancing its applicability in scenarios with limited computational resources.

Acknowledgement: None.

Funding Statement: This manuscript was prepared without any funding support.

Author Contributions: Wu-Jiu Sun wrote the main manuscript text and completed the experiment. Xiao Fan Liu provided theoretical ideas and revised the manuscript. All authors reviewed the results and approved the final version of the manuscript.

Availability of Data and Materials: In this study, we used the public datasets, which can be downloaded from the websites if needed (<https://archive.org/details/twitterstream> and <https://github.com/social-media-prediction/TPIC2017>, accessed on 01 October 2024.).

Ethics Approval: Not applicable.

Conflicts of Interest: The authors declare no conflicts of interest to report regarding the present study.

References

1. Wang S, Li C, Wang Z, Chen H, Zheng K. BPF++: a unified factorization model for predicting retweet behaviors. *Inf Sci.* 2020;515:218–32. doi:10.1016/j.ins.2019.12.017.

2. Liu Y, Chen Y, Fan ZP. Do social network crowds help fundraising campaigns? Effects of social influence on crowdfunding performance. *J Bus Res.* 2021;122:97–108. doi:10.1016/j.jbusres.2020.08.052.
3. Hayes JL, Brinson NH, Bott GJ, Moeller CM. The influence of consumer-brand relationship on the personalized advertising privacy calculus in social media. *J Interact Mark.* 2021;55(1):16–30. doi:10.1016/j.intmar.2021.01.001.
4. Wang J, Yang Y. Tweet retweet prediction based on deep multitask learning. *Neural Process Lett.* 2022;54:523–36. doi:10.1007/s11063-021-10642-3.
5. Wang L, Zhang Y, Yuan J, Hu K, Cao S. FEBDNN: fusion embedding-based deep neural network for user retweeting behavior prediction on social networks. *Neural Comput Appl.* 2022;34(16):13219–35. doi:10.1007/s00521-022-07174-9.
6. Sun WJ, Liu XF. Deep attention framework for retweet prediction enriched with causal inferences. *Appl Intell.* 2023;53(20):24293–313. doi:10.1007/s10489-023-04848-2.
7. Wei J, Tay Y, Bommasani R, Raffel C, Zoph B, Borgeaud S, et al. Emergent abilities of large language models. *Trans Mach Learn Res.* 2022. doi:10.48550/arXiv.2206.07682.
8. Jin B, Zhang W, Zhang Y, Meng Y, Zhang X, Zhu Q, et al. Patton: language model pretraining on text-rich networks. In: *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*; 2023; Toronto, ON, Canada. p. 7005–20.
9. Chowdhery A, Narang S, Devlin J, Bosma M, Mishra G, Roberts A, et al. PaLM: scaling language modeling with pathways. *J Mach Learn Res.* 2023;24(240):1–113.
10. Li J, He X, Wei L, Qian L, Zhu L, Xie L, et al. Fine-grained semantically aligned vision-language pre-training. *Adv Neural Inform Process Syst (NeurIPS).* 2022;35:7290–303.
11. Wang H, Feng S, He T, Tan Z, Han X, Tsvetkov Y. Can language models solve graph problems in natural language? *Adv Neural Inform Process Syst (NeurIPS).* 2024;36.
12. Zhang X, Bosselut A, Yasunaga M, Ren H, Liang P, Manning CD, et al. GreaseLM: graph REASoning enhanced language models for question answering. In: *International Conference on Representation Learning (ICLR)*; 2022.
13. Li X, Lian D, Lu Z, Bai J, Chen Z, Wang X. GraphAdapter: tuning vision-language models with dual knowledge graph. *Adv Neural Inform Process Syst (NeurIPS).* 2024;36.
14. Eldele E, Ragab M, Chen Z, Wu M, Kwok C, Li X. Contrastive domain adaptation for time-series via temporal mixup. *IEEE Trans Artif Intell.* 2023;1(1):1–10.
15. Park J, Lee J, Sohn K. Dual-path adaptation from image to video transformers. In: *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*; 2023; Vancouver, BC, Canada. p. 2203–13.
16. Houshy N, Giurghi A, Jastrzebski S, Morrone B, Laroussilhe QD, Gesmundo A, et al. Parameter-efficient transfer learning for NLP. In: *International Conference on Machine Learning*; 2019; Long Beach, CA, USA, PMLR. p. 2790–9.
17. Vaswani A, Shazeer N, Parmar N, Uszkoreit J, Jones L, Gomez AN, et al. Attention is all you need. *Adv Neural Inform Process Syst (NeurIPS).* 2017;30.
18. Devlin J, Chang MW, Lee K, Toutanova K. BERT: pre-training of deep bidirectional transformers for language understanding. In: *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies Volume 1 (Long and Short Papers)*; 2023; Minneapolis, MN, USA. p. 4171–86.
19. Radford A, Narasimhan K, Salimans T, Sutskever I. Improving language understanding by generative pre-training; 2017 [Cited 2024 Nov 09]. Available from: https://s3-us-west-2.amazonaws.com/openai-assets/research-covers/language-unsupervised/language_understanding_paper.pdf.
20. Hossain MR, Hoque MM, Dewan MAA, Hoque E, Siddique N. AuthorNet: leveraging attention-based early fusion of transformers for low-resource authorship attribution. *Expert Syst Appl.* 2025;262:125643. doi:10.1016/j.eswa.2024.125643.
21. Liu P, Zhang L, Gulla JA. Pre-train prompt and recommendation: a comprehensive survey of language modeling paradigm adaptations in recommender systems. *Trans Assoc Comput Linguist.* 2023;11:1553–71. doi:10.1162/tacl_a_00619.

22. Jiang Y, Liang R, Zhang J, Sun J, Liu Y, Qian Y. Network public opinion detection during the coronavirus pandemic: a short-text relational topic model. *ACM Trans Knowl Discov Data (TKDD)*. 2021;16(3):1–27.
23. Chen J, Dong H, Wang X, Feng F, Wang M, He X. Bias and debias in recommender system: a survey and future directions. *ACM Trans Inf Syst*. 2023;41(3):1–39. doi:10.1145/3564284.
24. Liu Y, Zhao J, Xiao Y. C-RBFNN: a user retweet behavior prediction method for hotspot topics based on improved RBF neural network. *Neurocomputing*. 2018;275:733–46. doi:10.1016/j.neucom.2017.09.015.
25. Zhao Z, Meng L, Xiao J, Yang M, Wu F, Cai D, et al. Attentional image retweet modeling via multi-faceted ranking network learning. In: *IJCAI'18: Proceedings of the 27th International Joint Conference on Artificial Intelligence*; 2018 Jul 13. p. 3184–90.
26. Yin H, Yang S, Song X, Li J. Deep fusion of multimodal features for social media retweet time prediction. *World Wide Web*. 2021;24(4):1027–44. doi:10.1007/s11280-020-00850-7.
27. Fatemi B, Halcrow J, Perozzi B. Talk like a graph: encoding graphs for large language models. In: *The Twelfth International Conference on Learning Representations (ICLR)*; 2023; Kigali, Rwanda.
28. Zaken E, Goldberg B, Ravfogel S. BitFit: simple parameter-efficient fine-tuning for transformer-based masked language-models. In: *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 2: Short Papers)*; 2022; Dublin, Ireland. p. 1–9.
29. Dosovitskiy A, Beyer L, Kolesnikov A, Weissenborn D, Zhai X, Unterthiner T, et al. An image is worth 16×16 Words: transformers for image recognition at scale. In: *International Conference on Learning Representations (ICLR)*; Vienna, Austria; 2021.
30. Xie E, Wang W, Yu Z, Anandkumar A, Alvarez J, Luo P. SegFormer: simple and efficient design for semantic segmentation with transformers. *Adv Neural Inform Process Syst (NeurIPS)*. 2021;34:12077–90.
31. Sun WJ, Liu XF, Shen F. Learning dynamic user interactions for online forum commenting prediction. In: *2021 IEEE International Conference on Data Mining (ICDM)*; 2021; Auckland, New Zealand; IEEE. p. 1342–7.
32. Bao K, Zhang J, Zhang Y, Wang W, Feng F, He X. TALLRec: an effective and efficient tuning framework to align large language model with recommendation. In: *Proceedings of the 17th ACM Conference on Recommender Systems (RecSys)*; 2023; Singapore. p. 1007–14.
33. Kim S, Kang H, Choi S, Kim D, Yang M, Park C. Large language models meet collaborative filtering: an efficient all-round LLM-based recommender system. In: *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (SIGKDD)*; 2024; Barcelona, Spain. p. 1395–406.
34. Loshchilov I, Hutter F. Decoupled weight decay regularization. In: *International Conference on Learning Representations*; 2018; Vancouver, BC, Canada.